Volume 221, number 2, 194–198 FEB 05075 September 1987

Variations with position of replication errors due to exonuclease warm-up

Philippe J. Lecomte and Jacques Ninio

Institut Jacques Monod, Tour 43, 2 Place Jussieu, 75251 Paris Cedex 05, France

Received 1 July 1987

A.A mismatch errors occurring during poly(dA) replication with the Klenow fragment of *E. coli* DNA polymerase I have been quantified. The A/T ratio measured for chains extended by 1–25 nucleotides decreases by a factor of at least 15 from beginning to end. The deduced true error rate may decrease by a factor of 2.5 at each successive nucleotide addition. When ddATP is used instead of dATP, the ddA/T ratio indicates little variation of the misincorporation probability with position. Thus, the accuracy improvement in the first case is due to a warm-up of the proofreading function.

DNA polymerase I; Processivity: Mismatch error; Proofreading; Mnemonic enzyme; *(E. coli)*

1. INTRODUCTION

In studies of the polymerization/excision ratio of *E. coli* DNA polymerase I, we showed that the proofreading function of this enzyme warmed up as the enzyme progressed along the template [1,2]. This implies that the misincorporation rate is highest at the beginning of a synthesis patch and then decreases. To quantify such a trend, we used the Klenow fragment of *PolI* to replicate poly(dA) with both a correct (T) and an incorrect (dA or ddA) nucleotide. The ratios of chain labelling with T and A show that errors decrease sharply with the number of elongation events, due to changes in proofreading efficiency. Furthermore, the length of the primer is shown to have a profound influence on the elongation process.

2. MATERIALS AND METHODS

Most procedures are described in [2]. The incubation mixtures contained 50 mM Tris-HCl buffer (pH 7.5), 75 mM KCl, 1.2 mM $MnCl_2$, 0.15 mg/ml BSA, 10 μg/ml inorganic pyrophosphatase, 50 μM (in nucleotides) $poly(dA)_{900}$ primed with oligo(dT) of length 10, 12, 15 or 20 at a ratio of 40 initiator molecules per template molecule, 30 μM ^{3}H- or ^{32}P-labelled TTP and, when present, 3 μM non-complementary nucleotide, either dATP or ddATP, cold or ^{32}P-labelled. The volume was 50 μl. The reactions were initiated by addition of the Klenow enzyme (1.7 U/ml) and, after 10 min incubation at 37°C, aborted by the addition of 10 μl of 0.1 M EDTA. Under these conditions, only a small percentage of the initiators are elongated, as determined by 5′ labelling.

2.1. *Densitometry*

For each experiment, several exposures of various durations were made so that each band was scanned in the optimal density range. The autoradiograms were scanned and quantitated with a Shimadzu CS-930 densitometer, using complete peak integration. Due to software misconception, all background corrections had to be recalculated.

Correspondence address: P.J. Lecomte, Institut Jacques Monod, Tour 43, 2 Place Jussieu, 75251 Paris Cedex 05, France

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

2.2. *Scintillation counting*

TTP was labelled with ^{3}H and dATP or ddATP with ^{32}P. After autoradiography, the gels were cut into 10 slices corresponding to nucleotide additions of lengths 1, 2, 3, 4, 5, 6–10, 11–20, 21–30, 31–40 and >40.

Both methods are reliable, but the lability of the exonuclease [3] introduces some variability. Thus, our results are presented as averages of 2–5 independent experiments.

3. RESULTS AND DISCUSSION

Let P be the length in nucleotides of the oligo(dT) primer. Consider a misincorporation occurring at the n-th elongation event, thus at position $P + n$. Assume the enzyme dissociates from the template after having incorporated a few more, say a nucleotides and let $m = n + a$. The misincorporation event at position n will be reflected as a radiolabelled chain of length $P + m$. More generally, misincorporations at position $P + n$ contribute somewhat to the labelling of all chains of lengths $\geqslant P + n$. Let e_m be the true error rate for the m-th elongation event and E_m be the measured, apparent error rate. E_m is the ratio of chains of the same length $P + m$, labelled with either the incorrect (A) or correct (T) nucleotide. In all rigour, the last nucleotide in a chain may be in a state of incomplete proofreading so that the terminal error rate f_m may be larger than the corresponding internal error rate e_m. Then

$$E_m = (e_1 + e_2 + \ldots + e_{m-1} + f_m)/m \quad (1)$$

If we write eqn 1 for elongation lengths m and $m - 1$, and form the quantity $s_m = mE_m - (m - 1)E_{m-1}$, we find:

$$s_m = mE_m - (m - 1)E_{m-1} = f_m - f_{m-1} + e_{m-1} \quad (2)$$

We assume here that internal and terminal error rates are equal, so that $s_m = e_m$. From our E_m measurements we would like to deduce how e_m varies with m.

In fig.1 we show the distribution of the products of replication of poly(dA) primed with initiators of lengths 10, 12, 15 and 20 labelled with either T or A. The very first bands for T are clearly visible for $P = 10$ or 12 (lanes B,E), but extremely faint for $P = 15$ or 20 (lanes H,K), suggesting that when polymerization starts at $P = 15$ or 20, there is a close to zero probability of dissociation for the next few nucleotide additions [4]. In fig.2, we show E_m, the A/T ratio, as a function of chain length, for the various primers. There is a sharp decrease of E_m with m, the number of added nucleotides. From $m = 1$ to $m = 10$, E_m is reduced by a factor of 5.5 for $P = 10$, 10 for $P = 12$, 14 for $P = 15$ and 18 for $P = 20$. Between $m = 10$ and $m = 25$, the A/T ratio decreases further by a factor of 3 for all initiator lengths. At $m = 25$, A/T is about 1% of the [dATP]/[TTP] input. All curves agree, within

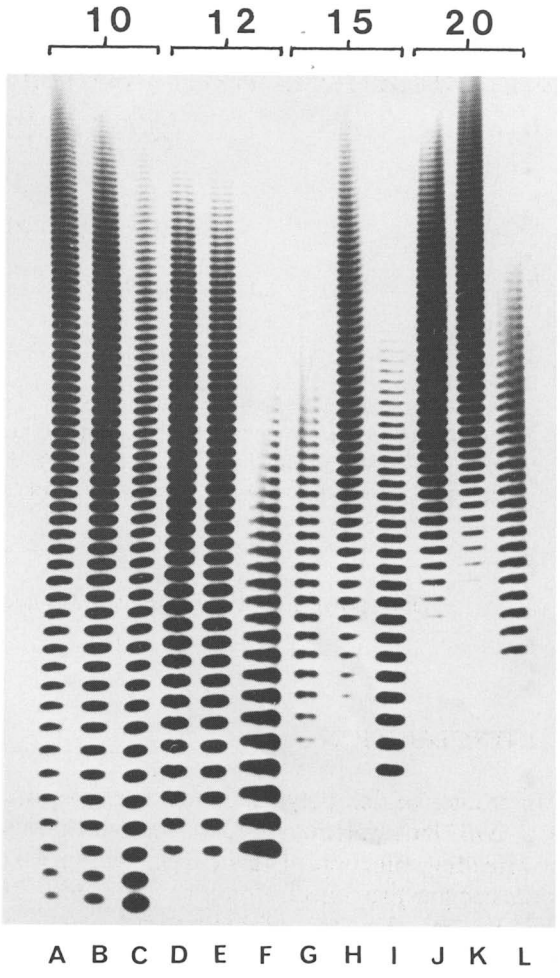

Fig.1. Replication of poly(dA)·oligo(dT) with TTP and dATP. For each primer length, 10, 12, 15 or 20, we show the gel-electrophoresis pattern of the chains elongated with TTP only (lanes A, D, G, J) or with both TTP and dATP, the label being on either T (B,E,H,K) or A (C,F,I,L).

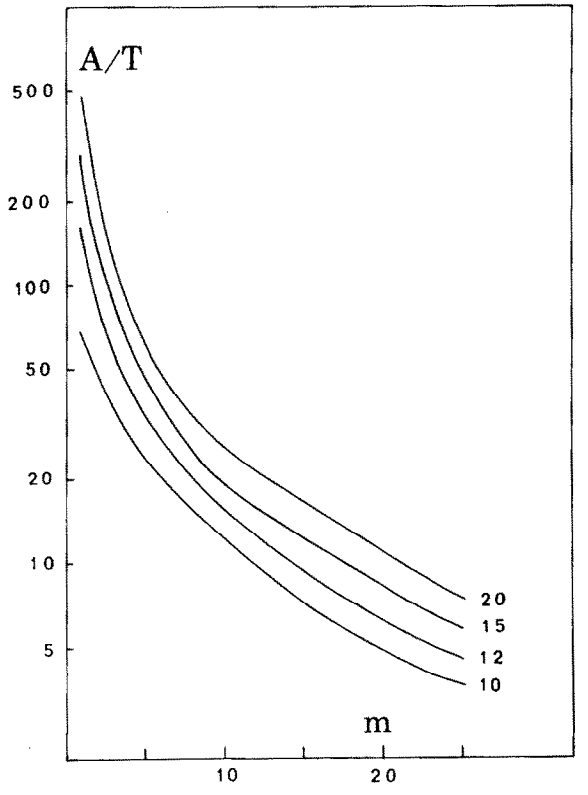

Fig.2. Misincorporation ratio, as a function of position. For primers of length $P = 10$, 12, 15 or 20, the ratio of chains labelled with A to those labelled with T is shown, on an arbitrary scale. Each curve represents the average of 3 independent densitometric scannings and incorporates the data from two further determinations by scintillation counting.

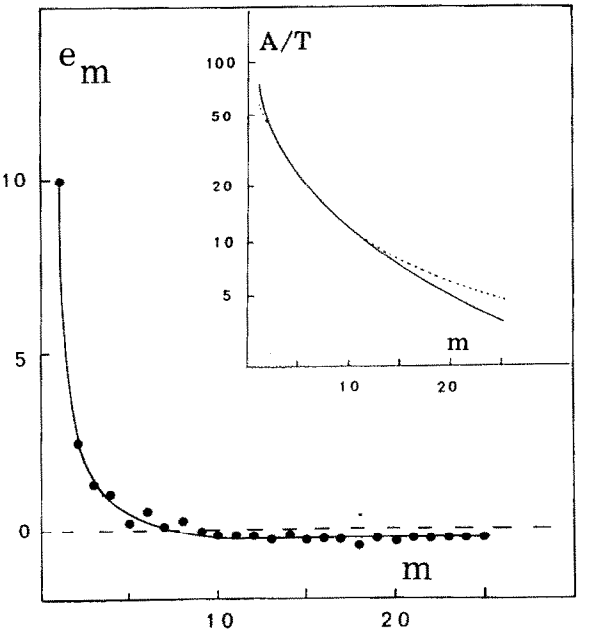

Fig.3. Elementary error rates. The results of fig.2 are condensed into the ideal A/T curve shown in the inset. From there, eqn 2 is applied to deduce the variations with position of elementary error rates (shown on an arbitrary scale). Taking the model $e_{m+1} = e_1 \times 0.4^m$, the interrupted A/T curve shown in the inset is predicted.

experimental error, beyond $m = 10$. The high A/T values, observed at small m for $P = 12$, 15 or 20 seem to originate from a deficit in T, rather than an excess of A (see fig.1). This deficit, arising from the presence of dATP (fig.1H,K), is less pronounced in the controls with TTP as sole substrate (lanes G,J). In fact, if the A/T ratio is computed taking lanes A, D, G and J as references instead of lanes B, E, H and K, this ratio is almost independent of initiator length for the first five points, and agrees with the normal ratio for $P = 10$. We may thus condense all experiments together into a single ideal curve. Starting with the curve for $P = 10$ we can introduce two slight corrections: for $m < 5$, we take the average A/T determined by comparison with lanes A, D, G and J. This shifts the position of the first point upwards slightly. For $m > 8$, we take an average of the four initial curves for $P = 10$, 12, 15 and 20. Again, the correction is minimal. Our working assumption is that this ideal curve (fig.3, inset) tells us how e_m varies with m. Consider a model in which the error rate follows a geometrical progression: $e_{m+1} = e_1 \times 0.4^m$. This model generates an E_m vs m curve (fig.3, inset) which approximates well the experimental curve at small m but deviates from it significantly at large m, due to the negative values of e_m (fig.3). Our provisional interpretation for the existence of negative s_m values is that an error at position m (slightly) increases the chances of abortion 10 or 20 nucleotides downstream. This leads in effect to an overestimate of A/T at small m and an underestimate at large m.

In order to discriminate between alternative interpretations, we performed polymerization reactions in which dATP was replaced by dideoxy-ATP (fig.4). Since incorporated ddNMPs are rather resistant to 3′-5′ excision [5], this ex-

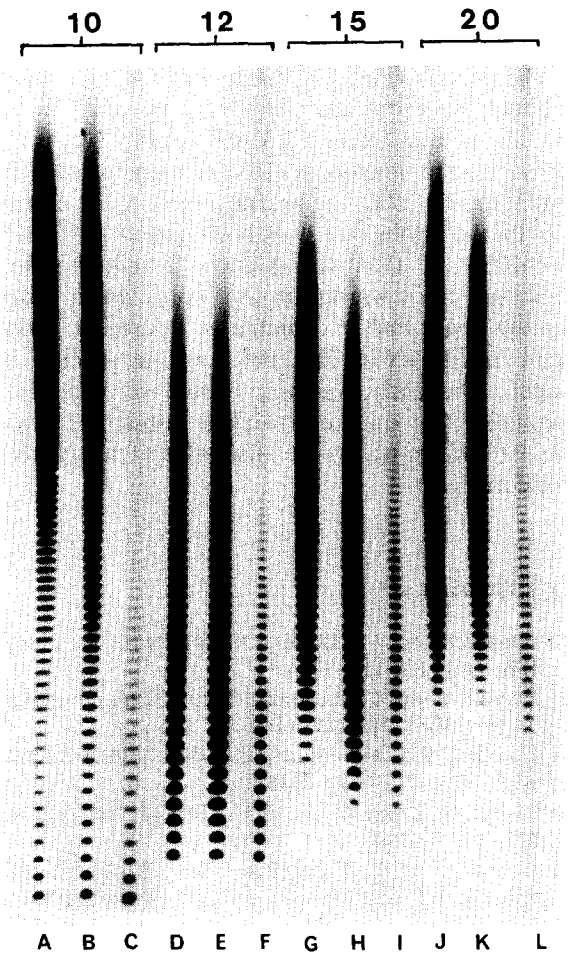

Fig.4. Replication of poly(dA)·oligo(dT) with TTP and dideoxy ATP. Same layout as in fig.1, ddATP replacing dATP.

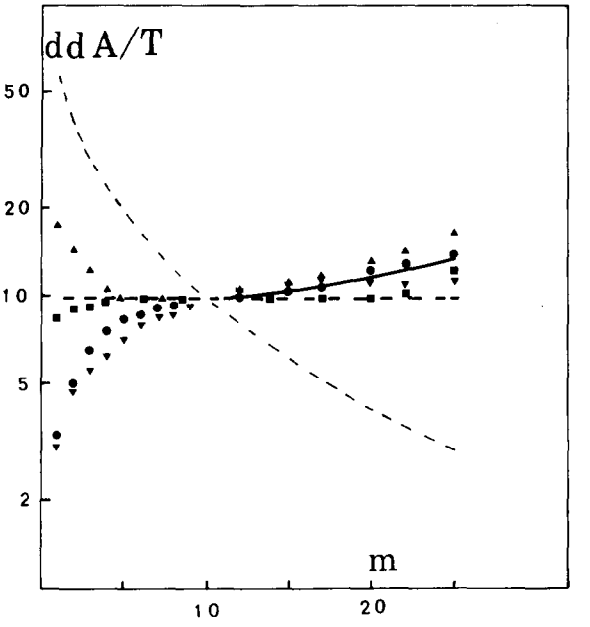

Fig.5. ddA termination frequency (arbitrary scale). Chains elongated by m nucleotides, terminated with ddA, are compared to the total of all chains elongated by m or more nucleotides (labelled with T). This requires division by m of the intensities of [^{32}P]TTP-labelled chains, and extrapolation to infinity. The data for P = 10 (▲), P = 12 (■), P = 15 (●) and P = 20 (▼) are averages of two independent experiments (one densitometry, one scintillation counting). The ideal A/T curve of the inset in fig.3 is shown for comparison, with an interrupted thin line.

perience should give us indications as to the misincorporation frequencies prior to proofreading.

Let us consider all chains of length $P + m - 1$ having bound an enzyme at the primer terminus. In the absence of ddATP, some of the chains will grow to various lengths: $P + q$, with $q \geqslant m$. Let R_m be the number of all such chains. In the presence of small amounts of ddATP, there will be some misincorporation of ddATP at position m producing a band of intensity r_m and causing termination of chains that would otherwise be susceptible to grow beyond that point to any length $> P + m$. Thus, the ddATP misincorporation at position m must be compared to R_m, i.e. the bulk of all chains of length $\geqslant P + m$. The true ddA/T error rate at position m is simply proportional to r_m/R_m, which is just the 'termination probability' introduced by McClure and Chow [6].

Let us analyze the results in terms of deviation from an ideal curve, the horizontal line, corresponding to a constant ddA/T ratio. This ratio varies little between m = 10 and 25, and the results for the four primers may be condensed into a single curve which departs from the horizontal by only 35% at m = 25. This deviation is the mirror image of the deviation observed between model and experiment in the case of A/T ratios (fig.3, inset). The standard misincorporation probability at m = 10 is about the same for all primers. Its value is around 0.05% of the [ddATP]/[TTP] input. For the first nucleotides, there is a primer-dependent systematic effect which again is the reverse of the

A/T trend. The longer the primer, the more important the deficit in ddA-terminated chains.

On the whole, variations in ddA/T are of a much smaller amplitude than those in A/T. For $P = 12$, ddA/T is a constant ± 20% from the 1st to the 25th nucleotide, whereas A/T varies by a factor > 30.

To sum up, the variations in A/T and ddA/T ratios seem to be due to two effects. A major effect, described by the ideal curves of fig.3 (inset) and fig.5 is a warm-up of the exonuclease, producing a rapid decline in the A/T ratio, but having no influence on ddA misincorporation. A minor effect, seen mainly for primers of lengths 15 and 20, would produce a deficit in the chains extended by a few nucleotides, provided they do not contain dA. Our probabilistic processivity model, reflected in eqns 1 and 2, assumes that the nucleotide at position $P + n$ has no influence on the probability of elongation to $P + n + 1$. The observed deviations at small m may perhaps reflect the inadequacy of this assumption which will require further testing.

Similar studies were conducted with intact *Pol*I (not shown). We also observed here a rapid decrease of A/T with position, for all initiator lengths. However, the analysis is more complex than with the Klenow fragment, because of a parasitic effect which needs to be elucidated. Everything occurs as if, with primers of lengths 15 or 20, there is an alternative pathway where the polymerase starts by excising processively 5–10 residues of the primer and only then starts to elongate.

The present results are compatible with a rapid exonuclease warm-up, since the true error rates, the e_m values, seem to drop rapidly with m. However, the mnemonic effect demonstrated with *Pol*I requires that the warm-up be more evenly distributed along the synthesis patch [2]. One possibility is that the Klenow fragment's exonuclease both warms up and cools down more rapidly than *Pol*I's exonuclease. Another more likely possibility is that the warm-up occurs in stages. The early warm-up would have an immediate effect on the erroneous substrate, while the late warm-up would also affect the correct nucleotide.

REFERENCES

[1] Papanicolaou, C., Dorizzi, M. and Ninio, J. (1984) Biochimie 66, 115–119.
[2] Papanicolaou, C., Lecomte, P. and Ninio, J. (1986) J. Mol. Biol. 189, 435–448.
[3] Lecomte, P.J. and Doubleday, O.P. (1983) Nucleic Acids Res. 11, 7505–7515.
[4] Detera, S.D. and Wilson, S.H. (1982) J. Biol. Chem. 257, 9770–9780.
[5] Atkinson, M.R., Deutscher, M.P., Kornberg, A., Russell, A.F. and Moffatt, J.G. (1969) Biochemistry 12, 4897–4904.
[6] McClure, W.R. and Chow, Y. (1980) Methods Enzymol. 64, 277–292.